BIO 244: Unit 7

Cox's Proportional Hazards Model

In this unit we introduce Cox's proportional hazards (Cox's PH) model, give a heuristic development of the partial likelihood function, and discuss adaptations to accommodate tied observations. We then explore some specific tests that arise from likelihood-based inferences based on the partial likelihood. Asymptotic properties of the resulting estimators and tests will be covered in later units.

7.1 Setting and PH Model: For each of n subjects we have the value of some covariate vector Z and the survival outcome (U, δ) representing noninformatively right-censored values of a survival time T. That is, for subject i, Z_i denotes the value of the covariate vector Z, and T_i and C_i denote the underlying survival time and potential censoring time, respectively, and we observe (Z_i, U_i, δ_i), where $U_i = min\{T_i, C_i\}$ and $\delta_i = 1[T_i \leq C_i]$, and where $T_i \perp C_i \mid Z_i$. The reasons why noninformative censoring is defined by the conditional independence of T_i and C_i, given Z_i, are discussed later.

One way to model a relationship between Z and T is by assuming $h(\cdot)$ is functionally related to Z.

e.g., $\quad T \sim Exp(\lambda_Z)$

$$\text{where} \quad h(t) = \lambda_Z = e^{\alpha+\beta Z} = \lambda_0 e^{\beta Z}$$

$$(\lambda_0 = e^{\alpha}).$$

Thus, we might assume that the T_i are independent with $T_i \sim Exp\left(\lambda_0 e^{\beta Z_i}\right)$, where Z_i = value of Z for subject i.

Note: $\beta = 0$ in the example means λ_Z does not depend on Z, and thus

that Z is not associated with T.

Let's Generalize: Let $h(t|Z)$ denote the h.f. for a subject with covariate Z.

Suppose that

$$h(t|Z) = \underbrace{h_0(t)}_{\substack{\text{function of } t, \\ \text{but not } Z}} \cdot \underbrace{g(Z)}_{\substack{\text{function of } Z, \\ \text{but not } t.}}$$

This is sometimes called a "multiplicative intensity model" or "multiplicative hazards model" or "proportional hazards model". This factorization implies that

$$\frac{h(t|Z = Z_1)}{h(t|Z = Z_2)} = \frac{g(Z_1)}{g(Z_2)} = \text{ independent of } t$$

$\longrightarrow$ "proportional hazards" (PH)! That is, the hazard ratio corresponding to any 2 values of Z is independent of time.

Important Special Case: $g(Z) = e^{\beta Z}$. This gives

$$h(t \mid Z) = h_0(t) \cdot e^{\beta Z} \tag{7.1}$$

$\Rightarrow$ **Cox's proportional hazards (Cox's PH) Model.**

Here

$$\frac{h(t|Z=Z_1)}{h(t|Z=Z_2)} = e^{\beta(Z_1-Z_2)}.$$

For scalar Z, e^{β} = hazard ratio corresponding to a unit change in Z.

In general, $\beta = 0 \Leftrightarrow Z$ not associated with T.

E.g., if

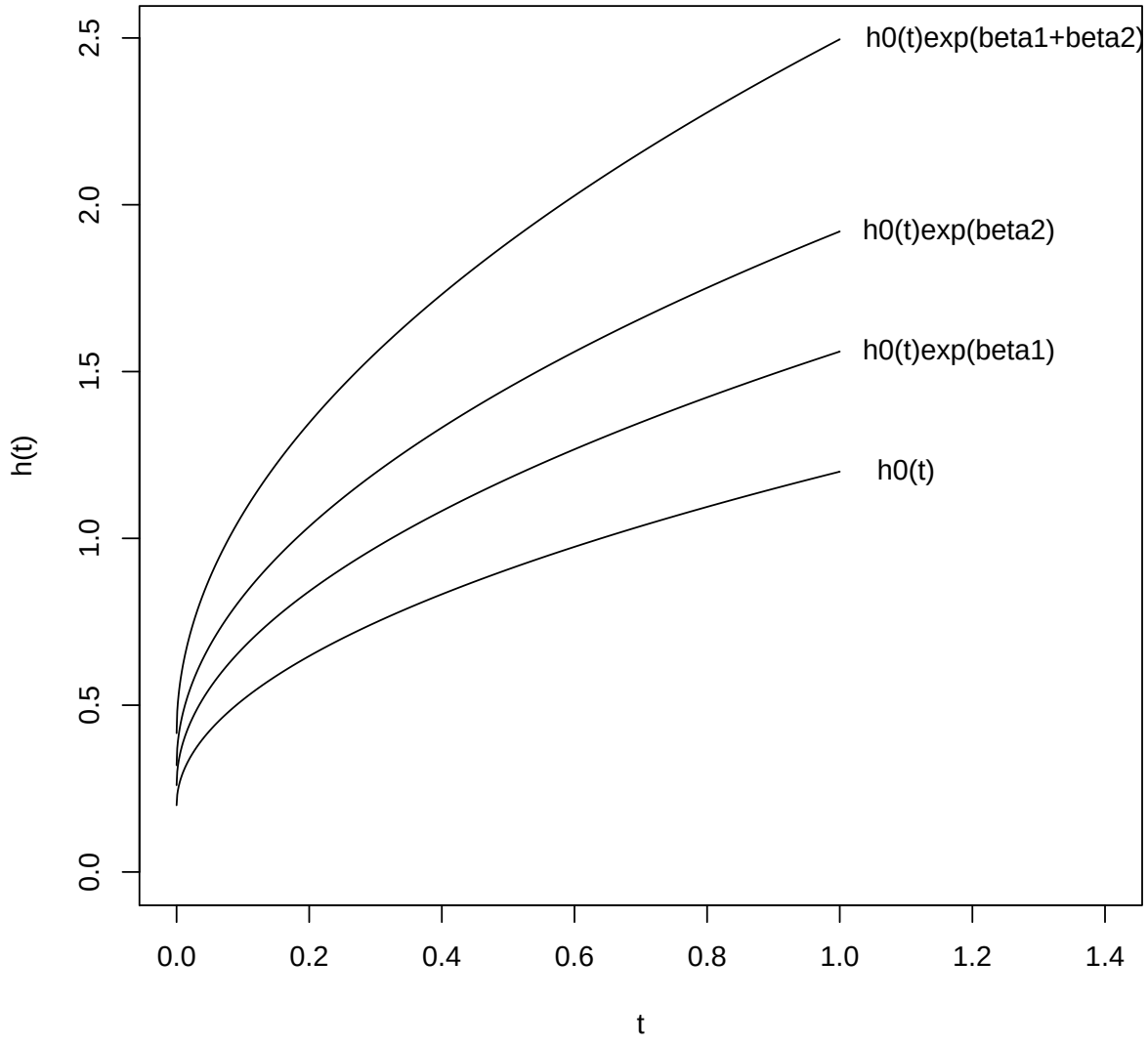
$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \qquad Z_1 = \begin{cases} 0 & \text{Rx (treatment)} \ 0 \\ 1 & \text{Rx (treatment)} \ 1 \end{cases}$$

$$Z_2 = \begin{cases} 0 & \text{female} \\ 1 & \text{male} \end{cases}$$

$$\text{and } \beta = (\beta_1, \beta_2),$$

then

$$h(t|Z) = \begin{cases} h_0(t) & \text{Rx } 0, \text{female} \\ h_0(t)e^{\beta_1} & \text{Rx } 1, \text{female} \\ h_0(t)e^{\beta_2} & \text{Rx } 0, \text{male} \\ h_0(t)e^{\beta_1+\beta_2} & \text{Rx } 1, \text{male} \end{cases}$$



7.2 Inference: How can we base inferences about β on (7.1)?

- Assume a parametric form for $h_0(t)$, conduct parametric analysis (e.g., $h_0(t) = \lambda_0$).

- Allow $h_0(t)$ to be arbitrary.

The latter is more general, but how do we carry out inference?

Note: (7.1) implies

$$S(u|Z) = (S_0(u))^{e^{\beta Z}},$$

$$\begin{aligned} \text{where} \quad S_0(u) &= e^{-\int_0^u h_0(t)dt} \\ &= \text{survival function for someone with Z} = 0 \\ &= S(u|0). \end{aligned}$$

Also, $f(u|Z) = h(u|Z)\ S(u|Z)$.

Thus, given n independent observations from (7.1), say (u_i, δ_i, z_i), the likelihood function is

$$\begin{aligned} L(\beta, h_0(\cdot)) &= \prod_i \Big(f(u_i|z_i)^{\delta_i}\ S(u_i|z_i)^{1-\delta_i} \Big) \\ &= \prod_i \Big(h(u_i|z_i)^{\delta_i}\ S(u_i|z_i) \Big) \\ &= \prod_i \left(\left(h_0(u_i)e^{\beta z_i}\right)^{\delta_i} \left(e^{-\int_0^{u_i} h_0(t)dt}\right)^{e^{\beta z_i}} \right) \\ &= \text{function} \Big(\text{data},\ \beta,\ h_0(\cdot) \Big). \end{aligned}$$

If we allow $h_0(\cdot)$ to be arbitrary, the "parameter space" is

$$\mathcal{H} \times \mathbb{R}^p = \left\{ (h_0(\cdot), \beta) \ :\ h_0(u) \geq 0 \text{ for all } u,\ \int_0^\infty h_0(u) = \infty \ \text{ and } \ \beta \in \mathbb{R}^p \right\},$$

where p is the dimension of the vector β. The condition

$$\int_0^\infty h_0(u)du = \infty$$

ensures that $S_0(\infty) = 0$.

Note: The primary goal in many applications is to make an inference about β, and the underlying hazard $h_0(\cdot)$ is a nuisance parameter (actually, a nuisance function).

Inferences in such settings are commonly called "semi-parametric". Standard likelihood theory, based on Euclidean parameter spaces, does not apply here, and thus other methods are needed.

Cox's Idea: Try to factor $L(\beta,\ h_0(\cdot))$ as

$$L(\beta,\ h_0(\cdot)) \ = \ \underbrace{L_1(\beta)}_{\substack{\text{function of } \beta, \text{ whose maximum } (\hat{\beta}) \text{ enjoys nice properties } (\hat{\beta} \xrightarrow{P} \beta) \ (\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{\mathcal{L}} \mathcal{N}) \text{ although perhaps inefficient}}} \ \cdot \ \underbrace{L_2(\beta,\ h_0)}_{\substack{\text{some function of } h_0(\cdot) \text{ and } \beta \text{ which contains relatively little information about } \beta.}}$$

Then, Cox tells us to base inferences about β on the partial likelihood function $L_1(\beta)$.

Aside: Aspects of this idea are not new. For example, with linear rank tests comparing two groups (and where censoring cannot occur), suppose

$X_1, \ldots, X_n$ independent r.v.'s

$Z_1, \ldots, Z_n$ indicators of group $Z_i = \begin{cases} 0 & \text{group 0} \\ 1 & \text{group 1} \end{cases}$

$$\begin{aligned} F(X_i \mid Z_i = 0) &= F_0(x) \\ F(X_i \mid Z_i = 1) &= F_1(x), \end{aligned}$$

$$H_0 \ : \ F_0(\cdot) \ = \ F_1(\cdot).$$

Note: $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Z} = (Z_1, \ldots, Z_n)$
are equivalent to knowing $\mathbf{Z}, \mathbf{r}$ ("which one"), and $\mathbf{X}_{(\ )}$ ("its value")
where $\mathbf{X}_{(\ )} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ and $\mathbf{r} = (r_1, \ldots, r_n)$,
where $r_i =$ rank of X_i.

Likelihood $= f(\mathbf{X}, \mathbf{Z}) = g(\mathbf{X}_{(\ )}, \mathbf{r}, \mathbf{Z}) = \underbrace{g_1(\mathbf{r}, \mathbf{Z})} \cdot g_2(\mathbf{X}_{(\ )} \mid \mathbf{r}, \mathbf{Z})$

↖ rank tests based on this

the Z and their

corresponding ranks.

$\Longrightarrow$ Since $(\mathbf{r}, \mathbf{Z})$ is a subset of the data, inferences based on $g_1(\cdot)$ will be valid (though possibly inefficient).

Back to Cox $\rightarrow$ Cox's idea is similar, but what he proposes for L_1 is not in general the pdf/pmf of a subset of the data, as above with rank tests.

$\therefore$ $L_1(\beta)$ called a **partial likelihood**.

- What is $L_1(\beta)$ and why is it intuitively reasonable?

  $\longrightarrow$ Assume there are no tied observations and no censoring.

Define

$$\tau_1 < \tau_2 < \cdots \quad = \quad \text{distinct times of failure}$$

$$R_j = \text{risk set at } \tau_j = \{\ell \mid U_\ell \geq \tau_j\},$$

and

$$Z_{(j)} = \text{value of } Z \text{ for the subject who fails at } \tau_j \ .$$

Note that knowledge of the τ_j, R_j, and $Z_{(j)}$ allows us to reconstruct the original data for this setting (recall we assume no censoring for now).

Then

$$L_1(\beta) \stackrel{def}{=} \prod_j \left\{ \frac{e^{\beta Z_{(j)}}}{\sum_{l \in R_j} e^{\beta Z_l}} \right\}. \tag{7.2}$$

Example: Z binary $n = 5$,

$(U_i, \delta_i, Z_i) = (16,1,1), (13,1,0), (21,1,1), (11,1,0), (12,1,1)$

$\tau_1, \ldots, \tau_5 = 11, 12, 13, 16, 21$

$R_1 = \{1,2,3,4,5\}, \quad R_2 = \{1,2,3,5\}, \quad R_3 = \{1,2,3\}$
$R_4 = \{1,3\}, \quad R_5 = \{3\}$

$Z_{(1)} = 0, \quad Z_{(2)} = 1, \quad Z_{(3)} = 0, \quad Z_{(4)} = 1, \quad Z_{(5)} = 1$

$$L_1(\beta) = \cdots = \left(\frac{1}{3e^\beta + 2}\right)\left(\frac{e^\beta}{3e^\beta + 1}\right)\left(\frac{1}{2e^\beta + 1}\right)\left(\frac{e^\beta}{2e^\beta}\right)\left(\frac{e^\beta}{e^\beta}\right)$$
$$= \text{function of } \beta.$$

7.3 Heuristic Justification of $L_1(\beta)$ and Modification to Allow Ties

Suppose first there are no tied or censored observations. Then the partial likelihood arises from two different arguments:

1. Conditioning Argument: Consider j^{th} term in (7.2)

$$e^{\beta Z_{(j)}} \Big/ \sum_{\ell \in R_j} e^{\beta Z_\ell} \, . \tag{7.3}$$

Conditional on surviving up to just prior to τ_j (being in risk set R_j), the probability of someone with covariate value Z failing at $t = \tau_j$ is

$$h_0(\tau_j) e^{\beta Z}.$$

Conditional upon R_j and the fact that someone fails at $t = \tau_j$, the probability that it is someone with covariate value Z^* ($Z^* = Z_\ell$ for some $\ell \in R_j$) is

$$\frac{h_0(\tau_j) e^{\beta Z^*}}{\sum_{\ell \in R_j} h_0(\tau_j) e^{\beta Z_\ell}} = \frac{e^{\beta Z^*}}{\sum_{\ell \in R_j} e^{\beta Z_\ell}}.$$

Thus, the contribution to L_1 from the observation that $Z_{(j)}$ is the covariate value of observed failure is (7.3).

The overall partial likelihood, L_1, is obtained by multiplying these contributions.

2. Rank Statistic Argument: When there are no ties or censoring, $(\mathbf{U},\ \boldsymbol{\delta},\ \mathbf{Z})$ is equivalent to $(\mathbf{U}_{(\ )}, \mathbf{r}, \mathbf{Z})$ and it can be shown that $L_1(\beta) =$ marginal distribution of $\mathbf{r}$ for given Z's (see appendix).

What if there is censoring?

It is easily accommodated. In fact, (7.2) also applies if there is censoring.

Without censoring, each successive R_j has one less element; with censoring, R_{j+1} can have ≥ 1 fewer elements than R_j. The rank statistic arguments become problematic, however, in the presence of censoring.

What about tied failure times?

For continuous T_i this happens with probability zero; but in the real world it is common.

$\longrightarrow$ Several ad-hoc modifications to (7.2) have been considered.

Most popular one (attributed to Breslow):

$\tau_1 < \tau_2 < \cdots < \tau_k$ distinct failure times.

$d_j = \#$ failures at τ_j.

$Z_{(j)}^{(1)},\ Z_{(j)}^{(2)},\ \ldots,\ Z_{(j)}^{(d_j)}$ = values of Z for the d_j subjects who fail at τ_j.

R_j = as before.

$$L_1(\beta) = \prod_{j=1}^{K} \left\{ \prod_{i=1}^{d_j} \left\{ \frac{e^{\beta Z_{(j)}^{(i)}}}{\sum_{\ell \in R_j} e^{\beta Z_\ell}} \right\} \right\}. \qquad (7.4)$$

Idea: treat the d_j failures at τ_j separately, using (7.2),
but use the same risk set for each.

7.4 Inferences based on Partial Likelihood: Above is the most common form of Cox's Partial Likelihood.

Idea: proceed as if this were the likelihood,

– maximizing value = $\hat{\beta}$ ("semiparametric MLE"; not obtainable in closed form)
– approximate variance of $(\hat{\beta})$ by inverse of observed information from L_1

$\rightarrow$ use Wald test, score test, LRT as in ordinary ML settings.

Note: When Z is scalar, it is easily verified that:

$$U(\beta) = \frac{\partial \ln L_1(\beta)}{\partial \beta} = \sum_{j=1}^{k} \left\{ \left(\sum_{i=1}^{d_j} Z_{(j)}^{(i)} \right) - d_j \bar{Z}_j(\beta) \right\}$$

where $$\bar{Z}_j(\beta) \stackrel{\text{def}}{=} \frac{\sum_{\ell \in R_j} Z_\ell e^{\beta Z_\ell}}{\sum_{\ell \in R_j} e^{\beta Z_\ell}}$$

$$= \text{weighted average of the } Z \text{ in } R_j^{\cdot}$$

$$\hat{I}(\beta) = \frac{-\partial^2 \ln L_1(\beta)}{\partial \beta^2} = \sum_{j=1}^{K} d_j \left(\frac{\sum_{\ell \in R_j} Z_\ell^2 e^{\beta Z_\ell}}{\sum_{\ell \in R_j} e^{\beta Z_\ell}} - \bar{Z}_j(\beta)^2 \right)$$

$$= \sum_{j=1}^{K} d_j \Big(\sum_{\ell \in R_j} \omega_\ell^{(j)} \left(Z_\ell - \bar{Z}_j(\beta) \right)^2 \Big),$$

where $$\omega_r^{(j)} = \frac{e^{\beta Z_r}}{\sum_{\ell \in R_j} e^{\beta Z_\ell}}.$$

From these we can do Wald, LRT, or score tests.

e.g.,

Wald based on $\hat{\beta} \overset{\text{apx}}{\approx} N\left(\beta, \hat{I}^{-1}(\hat{\beta})\right)$

Score Test of $H_0 : \beta = 0$: based on assuming $U(0)/\sqrt{I(0)} \approx N(0,1)$ under H_0:

$$U(0) = \cdots = \sum_{j=1}^{k} \left\{ \sum_{i=1}^{d_j} Z_{(j)}^{(i)} - d_j \cdot \bar{Z}_j(0) \right\}$$

$$\begin{aligned} \bar{Z}_j(0) &= \sum_{\ell \in R_j} Z_\ell \Big/ \sum_{\ell \in R_j} 1 \\ &= \text{average value of the } Z \text{ in } R_j. \end{aligned}$$

Special Case

Two-sample problem

$$Z_i = \begin{cases} 0 & \text{treatment } 0 \\ 1 & \text{treatment } 1. \end{cases}$$

Then

$$U(0) = \sum_j (O_j - E_j),$$

$$\text{where} \quad O_j = \sum_{i=1}^{d_j} Z_j^{(i)} = \#\text{ subjects in group 1 that fail at } \tau_j$$

$$E_j = d_j \cdot \bar{Z}_j(0) = d_j \cdot \tfrac{Y_1(\tau_j)}{Y(\tau_j)},$$

$$\text{where} \quad Y_1(\tau_j) = \#\text{ in group 1 at risk at } \tau_j$$
$$Y(\tau_j) = \#\text{ at risk at } \tau_j$$

$\longrightarrow$ same as numerator of logrank statistic!!!

The approximate variance of $U(0)$ is given by $I(0)$, and it can be shown that

$$\hat{I}(0) = \cdots = \sum_j V_j^*,$$

where $V_j^* = \dfrac{Y(\tau_j) - 1}{Y(\tau_j) - d_j} \cdot V_j$ $\quad (V_j = \text{logrank term}).$

Often, $\dfrac{Y(\tau_j) - 1}{Y(\tau_j) - d_j} \approx 1$ $\quad$ (exact if no ties).

Thus, logrank test can be viewed as arising from a Cox PH model as a score test. This connection is very important. It tells us that

- Logrank test is 'oriented' to PH alternatives
- Theoretical justification of its asymptotic distribution, provided we can show $L_1(\beta)$ has properties of real likelihood.

(More later.)

Exercises

1. Construct an example where $T \perp C \mid Z$, where Z is a covariate, but where T and C are not independent. This shows that informative censoring can be 'induced' from a noninformative setting by failing to control for a covariate.

2. Given $h\ (t|Z) = h_0(t)\ g(Z)$, find an expression for $S(t|Z)$ in terms of $g(\cdot)$ and $h_0(\cdot)$.

3. Suppose $h\ (t|Z) = h_0(t)\ e^{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3}$ where

$$Z_1 = \begin{cases} 0 & \text{tr. 0} \\ 1 & \text{tr. 1} \end{cases} \qquad Z_2 = \begin{cases} 0 & \text{female} \\ 1 & \text{male} \end{cases}$$

$$\text{and } Z_3 = Z_1 \cdot Z_2.$$

   What values of $\beta_1, \beta_2, \beta_3$ correspond to

   (a) treatment hazard ratio same in males as in females

   (b) no treatment effect in males, but an effect in females

   (c) no treatment effect

4. Verify the expression for $L_1(\beta)$ at the bottom of page 8.

5. Verify expressions for $U(\beta)$ and $\hat{I}(\beta)$ on page 11, and for $U(O)$ on page 12.

6. Consider a regression problem where X denotes treatment group (=0 or 1) and there is a single covariate Z. Suppose that the distribution of censoring does not depend on Z or on X, that the observations are noninformatively censored, and that $X \perp Z$. Let $p \stackrel{def}{=} P(X = 1)$ and assume that Z has the U(0,1) distribution.

(a) Derive expressions for the marginal hazard functions $h(t \mid X = 0)$ and $h(t \mid X = 1)$ in terms of the $f(t|X, Z)$.

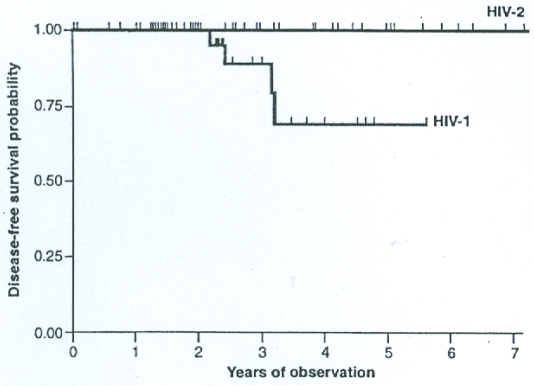
(b) Consider use of the ordinary logrank test to compare the treatment groups (ignoring any information about Z). Show that this is asymptotically a valid test of the hypothesis $H_0 : h(t \mid X = 0, Z) = h(t \mid X = 1, Z)$ for all t and Z, where $h(t \mid X, Z)$ denotes the hazard function at time t for someone in treatment group X and with covariate value Z. Make sure that you prove the conditions that you need for the logrank test to be valid.

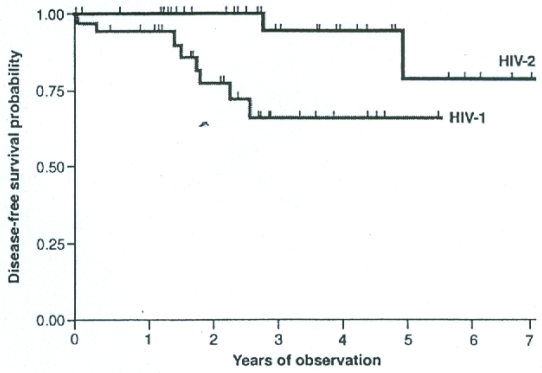
(c) Now suppose that $h(t \mid X, Z)$ is given by

$$h(t \mid X, Z) = h_0(t)e^{\alpha X+\beta Z} \ . \qquad (1)$$

An alternative test of H_0 can be obtained by the partial likelihood score test of $\alpha = 0$ based on fitting (1). Assuming (1) holds, how would you expect the efficiency of this test of H_0 to compare to that of the logrank test from part (b)? Give a heuristic justification for your answer.

7. The survival curves in this question are in an article in Science, 1994, volume 265, "Reduced rate of disease development after HIV-2 infection as compared to HIV-1", by Marlink et al, page 1589. HIV-2 is a close relative of the prototype AIDS virus, HIV-1. The article compares the prognosis of women with HIV-1 and HIV-2 in Dakar, Senegal. Of course, HIV-1 and HIV-2 infection has not been randomly assigned.





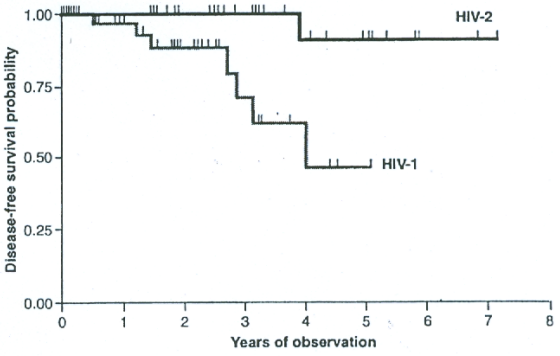


Fig. 1 (top, left). AIDS disease-free survival in seroincident women according to serostatus. AIDS as an outcome was defined according to the CDC revised surveillance definitions for AIDS (*13*). Kaplan-Meier disease-free survival analysis was performed on the seroincident women grouped by serostatus (*14*). Statistical significance was determined by Gehan's Wilcoxon test ($P = 0.02$) and by log-rank sum ($P = 0.01$). **Fig. 2 (top, right).** CDC IV disease-free survival in seroincident women, according to serostatus. CDC IV HIV-related disease as an outcome included all those conditions outlined in the CDC Classification System for HIV Infection fulfilling group IV disease criteria for HIV-related outcome, both AIDS defining and not AIDS defining (*16*). (Wilcoxon test, $P < 0.01$ and log-rank sum, $P = 0.01$). **Fig. 3 (bottom, right).** Disease-free survival as measured by abnormal $CD4^+$ count in seroincident women according to serostatus. Graph includes 31 of 32 HIV-1 seroincident women and 32 of 33 HIV-2 seroincident women. Disease outcome reflects a lymphocyte count of less than 400 cells/mm^3 (Wilcoxon test, $P < 0.01$ and log-rank sum, $P = 0.01$).

(a) What are the assumptions behind the Kaplan Meier estimator? And for the logrank test? How can results be interpreted? For this question, we are not interested in what tests are oriented towards but what are the fundamental assumptions and how can the results be interpreted (like: what exactly are you estimating and testing).

- (10 points). Can you come up with an example when these assumptions are not met, other than mentioned in the questions here?

(b) Suppose it is known that a measured confounding covariate X, e.g. categorized distance to a clinic, is not equally distributed among HIV-1 and HIV-2 infected women. What kind of analysis would you propose to *test* whether HIV-1 and HIV-2 have the same impact on AIDS free survival?

(c) Again, suppose it is known that a measured confounding covariate X, e.g. categorized distance to a clinic, is not equally distributed among HIV-1 and HIV-2 infected women. What kind of *model* would you propose for the hazard of failure? Propose two models and compare them.

(d) Suppose that 0 indicates the time these women first come to a clinic. If women with HIV-2 have fewer symptoms, they may show up later at the clinic for their first visit. How would that affect your interpretation of the figures on the previous page?

(e) ((Just a note: this kind of bias may not occur in the study this article is based on. Since February of 1985, all women registered as commercial sex workers at a specific hospital in Dakar, Senegal, have been serologically screened for exposure to HIV-1 and HIV-2 during biannual visits. The figures on the previous page are based on them.))

(f) Twelve women had both HIV-1 and HIV-2 infection. What happens to the properties of the logrank test if we put them in both groups?

(g) Eighty-five women moved from Dakar without health information follow up. How could that affect the estimates?

(h) Is it likely that Cox's proportional hazards model holds in the three figures in the paper? Describe this for each figure separately.

(i) Next to the figures in the paper, you can see the logrank test and the Wilcoxon test. What is the difference?

Appendix: Equivalence of $L_1(\beta)$ to marginal distribution of rank statistic when there are no ties or censored observations

This result was shown by Kalbfleisch & Prentice (1973). Suppose Z is a scalar. Let $\mathbf{r} = r_1, \ldots, r_n$ denote the rank statistic; i.e., $r_1 = 4$ means that subject 1 has the 4^{th} smallest failure time. Define $Z_{(1)}, \ldots, Z_{(n)}$ by $Z_{(r_i)} = Z_i$; i.e., $Z_{(i)}$ is the covariate value for the person with the i^{th} smallest failure time. The pdf of survival time for someone with covariate Z is given by

$$f\ (t|Z) = h_0(t)\ e^{\beta Z}\ \cdot\ \exp\left(-e^{\beta Z}\ H_0(t)\right).$$

The pmf of $\mathbf{r}$ for given constants $Z_1, Z_2, \ldots, Z_n$, $P(R = r|Z_1, \ldots, Z_n)$, can thus be expressed as

$$\begin{aligned} P(\mathbf{r}) &= \int_{\mathbb{R}^n : R=r} \prod_{j=1}^{n} f(t_j|Z_j)\ \ dt_1 \cdot dt_2 \ldots dt_n \\ &= \int_{\mathcal{S}} \prod_{j=1}^{n} f(t_j|Z_{(j)})\ \ dt_1 \cdot dt_2 \cdots dt_n, \end{aligned}$$

where

$$\mathcal{S} = \{(t_1, t_2, \ldots, t_n)\ :\ 0 \le t_1 < t_2 < \cdots < t_n\},$$

and $Z_{(r_j)} = Z_j$. Then

$$P(\mathbf{r}) = \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{n-1}}^\infty \prod_{j=1}^{n} \left\{ h_0(t_j)\ e^{\beta Z_{(j)}} e^{-e^{\beta Z_{(j)}} H_0(t_j)} \right\} \quad dt_n \cdot dt_{n-1} \cdots dt_1$$

$$= \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{n-2}}^\infty \prod_{j=1}^{n-1} \{\ \} \left[\int_{t_{n-1}}^\infty h_0(t_n)\ e^{\beta Z_{(n)}}\ \exp\left(-e^{\beta Z_{(n)}} H_0(t_n)\right) dt_n \right] dt_{n-1} \cdots dt_1$$

The term in square brackets equals

$$\exp\left\{-e^{\beta Z_{(n)}} H_0(t_{n-1})\right\};$$

i.e., the probability that someone with hazard function $h_0(u)e^{\beta Z_{(n)}}$ survives beyond t_{n-1}.

Thus, $P(\mathbf{r})$ becomes

$$\int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{n-2}}^\infty \prod_{j=1}^{n-1} \{\quad\} \cdot e^{-H_0(t_{n-1})\, e^{\beta Z_{(n)}}} \quad dt_{n-1} \cdots dt_2 dt_1$$

$$= \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{n-3}}^\infty \prod_{j=1}^{n-2} \{\quad\} \left[\int_{t_{n-2}}^\infty h_0(t_{n-1})\, e^{\beta Z_{(n-1)}}\, e^{-H_0(t_{n-1})\, e^{\beta Z_{(n-1)}}} \right.$$

$$\left. \cdot\, e^{-H_0(t_{n-1})\, e^{\beta Z_{(n)}}} \; dt_{n-1} \right] \quad dt_{n-2} \cdots dt_2$$

$$= \frac{e^{\beta Z_{(n-1)}}}{e^{\beta Z_{(n-1)}} + e^{\beta Z_{(n)}}} \cdot \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{n-3}}^\infty \prod_{j=1}^{n-2} \{\quad\} \cdot [\quad] \quad dt_{n-2} \cdots dt_2 dt_1,$$

where

$$[\quad] = \int_{t_{n-2}}^\infty h_0(t_{n-1}) \left(e^{\beta Z_{(n-1)}} + e^{\beta Z_{(n)}}\right) e^{-H_0(t_{n-1})\left[e^{\beta Z_{(n-1)}} + e^{\beta Z_{(n)}}\right]} \quad dt_{n-1}$$

$$= e^{-H_0(t_{n-2})\left[e^{\beta Z_{(n-1)}} + e^{\beta Z_{(n)}}\right]}.$$

Thus,

$$P(\mathbf{r}) = \left(\frac{e^{\beta Z_{(n-1)}}}{e^{\beta Z_{(n-1)}} + e^{\beta Z_{(n)}}} \right) \int_0^\infty \int_{t_1}^\infty \cdots \int_{t_{n-3}}^\infty \prod_{j=1}^{n-2} \{ \quad \} \cdot e^{-H_0(t_{n-2}) \left[e^{\beta Z_{(n-1)}} + e^{\beta Z_{(n)}} \right]}$$

$$dt_{n-2} \cdots dt_2 dt_1.$$

Continuing in this way, we see that

$$P(\mathbf{r}) \;=\; \cdots \;=\; \prod_{j=1}^{n-1} \left\{ \frac{e^{\beta Z_{(j)}}}{\sum_{\ell=j}^{n} e^{\beta Z_{(\ell)}}} \right\} \;=\; \prod_{j=1}^{n} \left\{ \frac{e^{\beta Z_{(j)}}}{\sum_{\ell \in R_j} e^{\beta Z_\ell}} \right\},$$

and hence that $P(\mathbf{r}) = L_1(\beta)$. Thus, when there is no censoring or tied data, Cox's partial likelihood is a "real" likelihood – i.e., the marginal density of the rank statistic $\mathbf{r}$.

References

Breslow, NE (1972). Contribution to the discussion of Cox (1972), Journal of the Royal Statistical Society B, 34: 216-217.

Cox DR (1972). Regression models and life tables (with Discussion). Journal of the Royal Statistical Society B, 34:187-220.

Kalbfleisch JD and Prentice RL (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika, 60: 267-278.